

Figure 7: Spatial distribution of Bias Scores at 2200 UTC for the 2.5 mm threshold.

void of stations, the  $BS_{st}$  will be low because forecasted precipitation does not spread over many stations. Conversely, if the model forecasts precipitation in an area with high density of stations, the  $BS_{st}$  increases. Since in different days precipitation occurs and is forecasted to occur in different areas, there will be a daily variation in scores associated with station distribution.

The uneven distribution of the stations should not, however, impact the scores attained for the whole forecasting period as long as, on average, rain is evenly distributed over the model domain. Correlations of rain deviation from the average with station density skew the BS. Additionally, correlations between station density and forecast quality also affect the BS.

To avoid problems caused by the uneven distribution of stations, verification scores may be presented as a spatial plot for a given time, without condensing all the stations in a single number (as in Figure 2). A spatial plot for the BS of the 16-hour precipitation forecast at the 2.5 mm threshold is shown in Figure 7. Although there is not a straight correspondence between the BS and topography or continentality, low BSs are generally present near the coast and high BSs about 100 km inland, which suggests that forecasted rain systems associated with the sea breeze penetrate too far inland before they start precipitating.

## 7. CONCLUSIONS

There are large differences in the values of the BS computed at the grid or at the stations. For the model configuration and verification dataset used in this

study, the BS computed at the stations is lower than the BS computed at the grid.

This discrepancy is due to the precipitation regime. During summer the rains are of convective nature, and precipitation events are isolated and have a large chance of occurring in between stations. Forecasted rain may fall in between stations, which leads to a small number of stations with forecasted precipitation, lowering the bias score computed at the stations.

The method used to interpolate model forecasted rain data to station location has a large impact on verification scores. Methods that cause the forecasted precipitation to spread over a large number of stations lead to an increased BS computed at the stations.

## 8. REFERENCES

- Barnes, S. L., 1973. Mesoscale objective map analysis using weighted time-series observations. NOAA Tech. Mem. ERL NSSL-62, U.S. Department of Commerce, 60 pp.
- Bernardet, L. R., 1999. Verification of high-resolution precipitation forecasts for the 1996 Atlanta Olympic Games. VIII Conference on Mesoscale Processes, June 28 - July 2, Boulder, Colorado, pp 365-368.
- Snook, J. S., P. A. Stamus, J. P. Edwards, Z. Christidis and J. A. McGinley, 1998. Local-domain analysis and forecast model support for the 1996 Centennial Olympic Games. *Wea. Forec.*, **13**, 138-150.
- Briggs, W. M. and R. Zaretski, 1998. The effect of randomly spaced observations on field forecast errors. 14th Conference on Probability and Statistics in Atmospheric Sciences, 11-16 January, Phoenix, AZ, pp 5-8.

## Acknowledgements

The author thanks Nita Fullerton for reviewing the manuscript and Edward Tollerud for making several suggestions to improve this work.

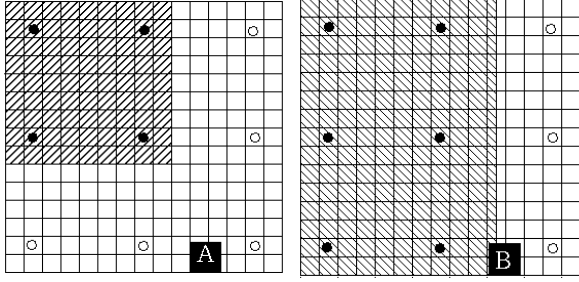


Figure 4: As Figure 3, but for a different rain event.

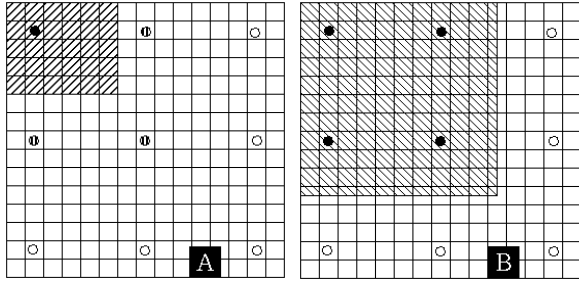


Figure 5: As Figure 3, but for a different rain event. The half filled dots in a) are gauges that receive interpolated forecasted precipitation when the 6x6 method is used.

situations may be envisioned in which  $BS_{grid} = BS_{2x2}$ , even if  $BS \neq 1$ .

We now proceed to investigate situations in which  $BS_{grid} \neq BS_{2x2}$ . Figures 4a and 4b show a situation in which  $BS_{2x2} > BS_{grid}$ . The model forecasted precipitation in  $10 \times 10 = 100$  grid points, or 4 stations. The verification data have observed rain in 6 stations, or  $11 \times 16 = 176$  grid points. In this case,  $BS_{grid} = 0.57$  and  $BS_{2x2} = 0.66$ , so  $BS_{grid} < BS_{2x2}$ .

The opposite situation is seen in Figures 5a and 5b. The model forecasted precipitation in  $7 \times 6 = 42$  grid points, or 1 station. The observations show precipitation in 4 stations or  $11 \times 11 = 121$  grid points, yielding  $BS_{grid} = 0.35$  and  $BS_{2x2} = 0.25$ , so  $BS_{grid} > BS_{2x2}$ .

These simple examples show that the relative magnitudes of the  $BS_{2x2}$  and the  $BS_{grid}$  are determined by the spread of the model forecasted precipitation over the stations. In Figure 5 the forecasted precipitation is widespread when compared to the observations, while in Figure 4 it is *not*.

To support this idea, the  $BS_{6x6}$  was computed. In Figure 5a, the model forecasted precipitation using the 6x6 method spreads over four stations, instead of just one for the 2x2 method. In this case,  $BS_{6x6} = 4/4 = 1$ , so  $BS_{grid} < BS_{6x6}$ .

Therefore, the type of interpolation used to bring the model data to the stations has large impact in

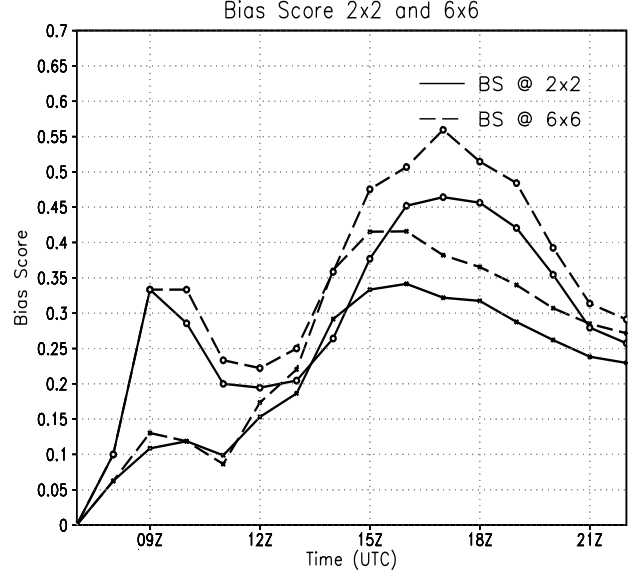


Figure 6: Time series of the BS for the 2.5 (xs) and 6.25 mm (circles) thresholds. Results are presented for the verification at the stations using the 2x2 method and the 6x6 method.

the scores attained. In particular, how many model grid points are used to compute the forecasted precipitation in a station is important in determining the spread of the forecasted data over the stations. Of course the radius of influence of the observed precipitation, and how many grid points each station influences is also very important. Briggs and Zaretzki (1998) used different interpolation techniques to grid geopotential data from idealized station observations and showed that the verification scores obtained were highly dependent on the method used.

## 6.2 Applications to this dataset

The situation depicted in Figure 5, in which,  $BS_{grid} > BS_{2x2}$  resembles the results showed in Figure 2. The model produces localized rain events at the 2.5 mm and higher thresholds. The events cover mainly the area in between stations and may not encompass a significant number of stations.

The use of a 6x6 scheme for the set of all forecasts showed that the  $BS_{6x6}$  is larger than the  $BS_{2x2}$  (Figure 6).

## 6.3 The impact of unevenly spaced stations

As can be seen in Figure 1, the stations are not evenly distributed over the model domain. This has a large impact on the scores on individual days. If in a given day the model forecasts precipitation in an area

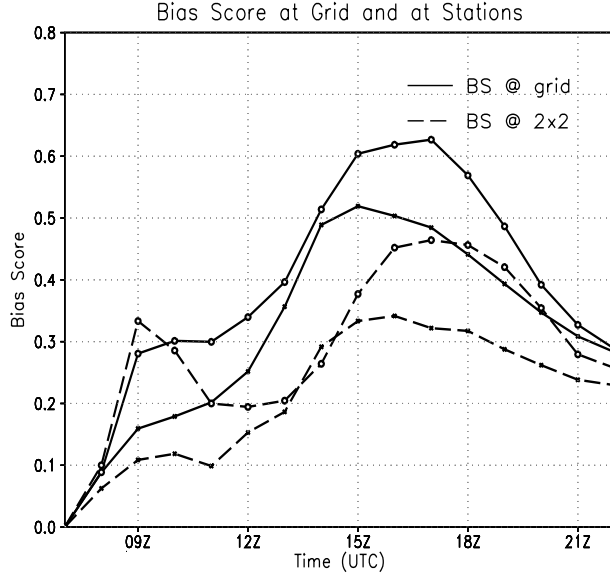


Figure 2: Time series of the BS for the 2.5 (xs) and 6.25 (circles) thresholds. Results are presented for the verification at the grid and at the stations, using the 2x2 method.

percent of the amplitude was retained for waves of 120–km wavelength and a smaller retaining response was used for shorter waves.

#### 4.2 Algorithm to compute the BS at the stations

$BS_{st} = (\# \text{ of stations with forecasted precipitation at or above a threshold}) / (\# \text{ of stations with observed precipitation at or above a threshold})$ .

To obtain (# of stations with forecasted precipitation at or above a threshold) the modeled data was interpolated to station location. Two types of bi-linear interpolation were used.

- Using the 4 points surrounding a station. This method is referred to as  $BS_{2 \times 2}$ . Since the model grid spacing is 8 km, the radius of influence of a forecasted event is 8 km.
- Using the 36 points surrounding a stations. This method is referred to as  $BS_{6 \times 6}$ . Since the model grid spacing is 8 km, the radius of influence of a forecasted event is 24 km.

## 5. RESULTS

Figure 2 shows the time series of the BS for the 2.5 and 6.25 mm thresholds. Results are presented for the verification at the grid and at the stations, using the 2x2 method.

Several features can be noted:

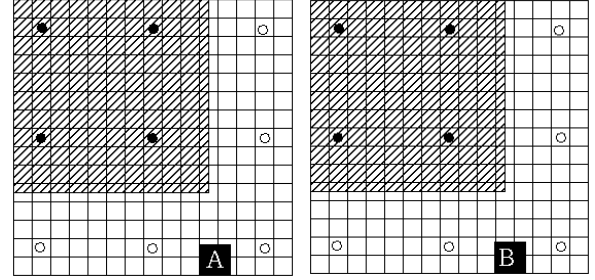


Figure 3: Idealized model grid and gauge locations. a) The shaded region represents the area covered by forecasted precipitation. The filled (open) dots are gauges with (without) interpolated forecasted precipitation. b) The filled (open) dots represent gauges with (without) observed precipitation. The shaded region represents the area covered by observed precipitation analyzed to the model grid.

1. At earlier times the BS is low, since the model does not have clouds or rain at the initial time;

2. The BS usually increases in the first 10 hours of forecast. This means that the model is increasingly producing clouds and precipitation. This can be considered the model spin-up time.

3. The BS reaches a maximum around 1500 UTC (9:00 AM EST), after which it decreases. The reduction (not shown, see Bernardet 1999) is related to the inability of the model to keep up with the observed development of afternoon precipitation.

4. The BS is less than one for all thresholds and times; therefore, the model is consistently under-forecasting precipitation.

5. For the 2.5 and 6.25 mm thresholds (and all higher thresholds — not shown) the  $BS_{grid}$  is larger than the  $BS_{st}$ .

It is this last result that concerns us in this paper. We will try to understand what causes the difference between the  $BS_{grid}$  and the  $BS_{st}$ .

## 6. DISCUSSION

### 6.1 Simplified System

To make this complex problem easier to understand, we will simplify it by assuming a domain of  $16 \times 16 = 256$  grid points with 9 rain gauges.

To start we assume a situation in which the model forecasted precipitation in  $11 \times 11 = 121$  grid points. When this forecast is interpolated to the stations using the 2x2 method, 4 stations are forecasted to have precipitation (Figure 3a). The verification dataset for that day has rain in 4 stations, and the objective analysis spreads the observed rain over 121 grid points (Figure 3b). In this case  $BS_{grid} = BS_{2 \times 2} = 1$ . Similar

## QPF VERIFICATION AT THE MODEL GRID VERSUS AT THE STATIONS

Lígia R. Bernardet\*  
 NOAA Forecast Systems Laboratory  
 Boulder, Colorado

## 1. INTRODUCTION

Scores commonly used to verify model quantitative precipitation forecast (QPF), such as the Bias Score (BS), require that forecasted and observed precipitation be at the same location. Therefore, either the observed values must be brought to the model's grid, or the forecasted precipitation must be interpolated to the station locations. Although these analysis procedures have a large impact on the final scores obtained for verification, they are seldom discussed in the literature of forecast verification.

In this paper, verification of high-resolution precipitation forecasts produced by a local model for the 1996 Atlanta Olympic Games will be presented. A comparison will be made between verification at the stations and verification at the model grid. The results suggest that the differences in the scores depend on the density of stations compared to the density of model grid points, on the heterogeneity of station distribution, on the the radius of influence of each rain event in the analysis algorithm, and on the precipitation regime.

## 2. MODEL SETUP

The Scalable Forecast Model (SFM) was configured with one grid with 8-km horizontal grid spacing and 85 points in each horizontal direction, covering the state of Georgia and parts of Tennessee, North Carolina, South Carolina, Florida, and Alabama (Snook et al., 1998). The vertical grid spacing was 100 m near the surface, stretching to 1000 m aloft. Forecasts were initialized using the Local Analysis and Prediction System (LAPS; Snook et al., 1998) at 0600 UTC, or 0200 local time, and run for 16 hours. Forecasts made with the 29-km Eta model were used as boundaries toward which the SFM was nudged.

Statistics were computed for a set of 19 days between 18 July and 24 August, 1996.

---

\*Also affiliated with the National Research Council. Corresponding Author Address: Dr. Lígia R. Bernardet; NOAA FSL; 325 Broadway R/E/FS1; Boulder, CO 80303-3328; e-mail: ligia@fsl.noaa.gov.

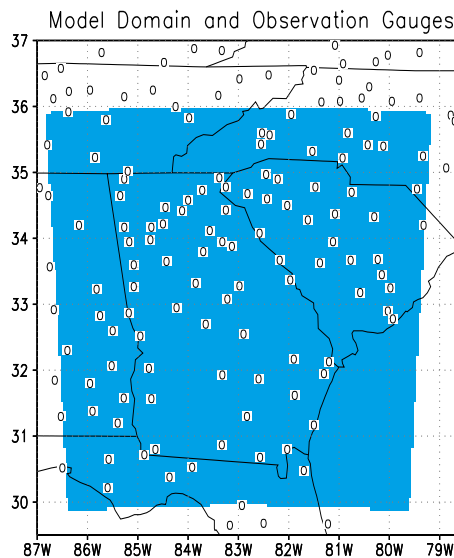


Figure 1: Model domain and location of observation gauges.

## 3. DATA

The precipitation data used in this study were obtained from the Hourly Precipitation Dataset, administered by the National Oceanic and Atmospheric Administration's National Climatic Data Center. One hundred and twenty rain gauges were used for verification.

Figure 1 shows the model domain and the location of the rain gauges used for verification. The average spacing between stations is about 60 km.

## 4. SCORES

4.1 Algorithm to compute the BS at the grid

$BS_{grid} = (\# \text{ grid points with forecasted precipitation at or above a threshold}) / (\# \text{ grid points with observed precipitation at or above that threshold})$ .

To obtain (# grid points with observed precipitation at or above that threshold) a Barnes (1973) analysis was applied to the station observations. Ninety